On Well Quasi Orders of Free Monoids

Lucian Ilie
Arto Salomaa

Turku Centre for Computer Science
TUCS Technical Report No 95
February 1997
ISBN 951-650-965-7
ISSN 1239-1891

Abstract

The paper investigates down-sets associated to well quasi orders. Of particular language-theoretic interest is the quasi order $u \leq_s v$ (resp. $u \leq_{\mathcal{P}} v$) of u being a subword (resp. a Parikh subword) of v, as well as their inverses. We establish a number of results about the regularity and effective regularity of the down-sets, in particular, a general condition for the effective regularity of the down-set associated to an inverse well quasi order (Theorem 7.1). It is decidable whether or not an arbitrary regular language results as a down-set from an infinite or bi-infinite word, whereas the same problem is undecidable for context-free language. A quasi order being a well quasi order is connected to arbitrary languages being confluent.

1 Introduction

The results of Higman, [Hi], and Haines, [Ha], are often used in the theory of formal languages. For instance, let $u \leq_s v$ denote that u is a (scattered) subword of v. Then any language L whatsoever contains only a finite number of words minimal with respect to the partial order $\leq_s$. Consequently, if we have a set S of words such that no two words in S are comparable with respect to $\leq_s$, then necessarily S is finite.

Based on similar ideas, Haines also showed that the languages

$$\begin{aligned}&\textsc{down}(L) = \{w \mid w \leq_s u \text{ for some } u \in L\} \text{ and}\\ &\textsc{up}(L) = \{w \mid u \leq_s w \text{ for some } u \in L\}\end{aligned}$$

are regular for any language L. (This result is reported in Theorem 3.3 below.) Of course, the languages mentioned are effectively regular only in exceptional cases.

We investigate in this paper *down-sets*, defined analogously with the set $\textsc{down}(L)$ above. However, our starting point will be an arbitrary *quasi order* $\leq$, instead of $\leq_s$. We have chosen a quasi order as our starting point, rather than an arbitrary partial order, since our main language-theoretic applications concern the relation $u \leq_{\mathcal{P}} v$, u is a Parikh subword of v. This relation is not a partial order. We will mainly investigate the problem of the *regularity* of a down-set. The problem is related to the property of *confluence*: a language L is *confluent* with respect to a quasi order $\leq$ if and only if, for any $x, y \in L$, there is a $z \in L$ such that $x \leq z$ and $y \leq z$. For general theoretical results, it is important that our basic quasi order $\leq$ is a *well quasi order*, a requirement certainly satisfied by the relation $\leq_{\mathcal{P}}$.

After the fundamental definitions and a survey of previous results, given in sections 2 and 3, we investigate subwords of infinite, or bi-infinite, words in section 4. Specifically, we try to represent a down-set as the set of prefixes, suffixes or factors of a single infinite or bi-infinite word, when the underlying relation is $\leq_s$. When is such a representation possible? We give a decision method for regular languages, and show the problem to be undecidable for context-free languages.

Section 5 deals with the quasi order $\leq_{\mathcal{P}}$ and, supplemented with section 7, contains our most important contributions to the theory of formal languages. We show that the down-set of an arbitrary regular language, and even of an arbitrary context-free language is effectively regular. This result holds true also if we start with the inverse relation $\leq_{\mathcal{P}}^{-1}$. In fact, we prove a general result concerning the effective regularity of down-sets with respect to inverses of well quasi orders with certain properties. Our result applies to the family of context-free languages for both quasi orders $\leq_s$ and $\leq_{\mathcal{P}}$. It is also decidable whether or not an arbitrary context-free language is confluent with respect to $\leq_{\mathcal{P}}$, a result we consider to be unexpected. Our general result, Theorem 7.1, should be of interest independently of its applications to regular and context-free languages.

Finally, the notion of confluence is generalized in section 6. We show that this generalization is closely connected with the concept of well quasi order.

We feel that our algebraic approach is, in general, in accordance with the work of Marco Schützenberger. We would like to add the following specific detail. It is quite remarkable that the particular types of regular languages coming up in our considerations in sections 5 and 6 were investigated already in the 60's. Indeed, they were called "quasi-uniform events" in [Ro] and represent low loop complexity from the point of view of the accepting automaton. The well-known paper [DS] presents in a very compact fashion (a more elaborated treatment is given in [Sa2]) the whole landscape of loop complexities.

2 Definitions

For an alphabet Σ, we denote by Σ^* the free monoid generated by Σ and by λ, the empty word, its identity. The free semigroup generated by Σ is denoted by Σ^+.

A *right-infinite* word over Σ is a word which is unbounded from the right. It can be viewed as a function $\alpha : \mathbf{Z}_+ \longrightarrow \Sigma$ from the set of positive integers into the alphabet Σ. Analogously, a *left-infinite* word is a function $\alpha : \mathbf{Z}_- \longrightarrow \Sigma$. The set of all right-(left-)infinite words is denoted by Σ^ω (${}^\omega\Sigma$). For a finite word $w \in \Sigma^*$, $|w|$ denotes the length of w and, for $a \in \Sigma$, $|w|_a$ denotes the number of occurrences of a in w. Denote also $w^\omega = www\ldots \in \Sigma^\omega$ and ${}^\omega w = \ldots www \in {}^\omega\Sigma$.

A *bi-infinite* (two-sided) word is an infinite word without any end. We can define a bi-infinite word as a function $\alpha : \mathbf{Z} \longrightarrow \Sigma$ or, in fact, as an equivalence class of the set $\Sigma^{\mathbf{Z}}$ with respect to the equivalence relation defined for $\alpha, \beta \in \Sigma^{\mathbf{Z}}$ by $\alpha \sim \beta$ if and only if there is an integer k such that for any $n \in \mathbf{Z}, \alpha(n) = \beta(n+k)$. We denote by ${}^\omega\Sigma^\omega$ the set of all bi-infinite words over Σ.

Given a quasi order relation $\leq$ on Σ^*, we say that a language $L \subseteq \Sigma^*$ is *confluent w.r.t.* $\leq$ if and only if for any $x, y \in L$ there is $z \in L$ such that $x \leq z$ and $y \leq z$. Notice that the empty language is confluent w.r.t. any quasi order on Σ^*. Also, for any two quasi orders $\leq_1$ and $\leq_2$ on Σ^*, if $\leq_1 \subseteq \leq_2$, then any language confluent w.r.t. $\leq_1$ is confluent also w.r.t. $\leq_2$.

Some of the most important quasi orders on Σ^* in combinatorics on words, see [CK], are as follows, where we suppose that u, v are arbitrary finite words over Σ:

prefix:	$u \le_p v$ iff there is $w \in \Sigma^*$ such that $v = uw$,
factor:	$u \le_f v$ iff there is $w, z \in \Sigma^*$ such that $v = wuz$,
subword:	$u \le_s v$ iff $u = a_1 a_2 \ldots a_n, n \ge 0, a_i \in \Sigma, 1 \le i \le n$, $v = v_1 a_1 v_2 a_2 \ldots v_n a_n v_{n+1}, v_i \in \Sigma^*, 1 \le i \le n+1$,
Parikh subword:	$u \le_{\mathcal{P}} v$ iff $u = a_1 a_2 \ldots a_n, n \ge 0, a_i \in \Sigma, 1 \le i \le n$, and, for some $\pi \in \mathcal{S}_n, a_{\pi(1)} a_{\pi(2)} \ldots a_{\pi(n)} \le_s v$.

Notice that $\le_p, \le_f$, and $\le_s$ are partial orders but $\le_{\mathcal{P}}$ is not.

The *down-set of L w.r.t.* $\le$, denoted $\downarrow_{\le} L$, is the set

$$\downarrow_{\le} L = \{w \in \Sigma^* \mid w \le u \text{ for some } u \in L\}.$$

For instance, $\downarrow_{\le_f} L$ is the set of all factors of words in L, $\downarrow_{\le_p} L$ is the set of all prefixes of words in L. The down-operator $\downarrow_{\le}$ is monotone and idempotent.

For an infinite word $\alpha \in \Sigma^\omega \cup {}^\omega\Sigma \cup {}^\omega\Sigma^\omega$, we denote by $Fact(\alpha)$, $Pref(\alpha)$, $Suf(\alpha)$, the set of all finite factors, prefixes, and suffixes of α, respectively. For instance, for $\alpha \in^\omega \Sigma$, $Pref(\alpha) = \emptyset$.

For a quasi order $\le$ on Σ^* we define the following families of languages:

$$\begin{aligned}
\mathcal{F}^r_{\le} &= \{L \subseteq \Sigma^* | L = \downarrow_{\le} Pref(\alpha), \text{ for some } \alpha \in \Sigma^\omega\},\\
\mathcal{F}^l_{\le} &= \{L \subseteq \Sigma^* | L = \downarrow_{\le} Suf(\alpha), \text{ for some } \alpha \in {}^\omega\Sigma\},\\
\mathcal{F}^{bi}_{\le} &= \{L \subseteq \Sigma^* | L = \downarrow_{\le} Fact(\alpha), \text{ for some } \alpha \in {}^\omega\Sigma^\omega\},\\
\mathcal{F}_{\le} &= \mathcal{F}^r_{\le} \cup \mathcal{F}^l_{\le} \cup \mathcal{F}^{bi}_{\le}.
\end{aligned}$$

Examples

1. The context-free language

$$L_1 = \{a^n b^n \mid n \ge 0\}$$

is confluent w.r.t. $\le_f$ because, for any $a^n b^n, a^m b^m \in L_1$, we have, for any $p \ge \max\{m, n\}$, $a^n b^n \le_f a^p b^p, a^m b^m \le_f a^p b^p$, and $a^p b^p \in L_1$. It follows that L_1 is confluent w.r.t. $\le_s$ too. Since there is no word in L_1 which has as prefixes both ab and $a^2 b^2$, L_1 is not confluent w.r.t. $\le_p$. Notice that also Σ^* fails to be confluent w.r.t. $\le_p$, for any alphabet $\Sigma, \mathrm{card}(\Sigma) \ge 2$.

2. The regular language

$$L_2 = a^* \cup b^*$$

is confluent w.r.t. none of the partial orders $\le_f, \le_s$, or $\le_p$ since, for any $w_1 \in a^+, w_2 \in b^+$, there is no word in L_2 which contains both w_1 and w_2 as factors, subwords, or prefixes, respectively.

3. The language

$$L_3 = a^*b^*c^* \cup \{abca\}$$

is confluent w.r.t. $\leq_\mathcal{P}$ since $a^*b^*c^*$ is confluent w.r.t. $\leq_s$ and $abca \leq_\mathcal{P} a^2bc$. On the other hand, L_3 is not confluent w.r.t. $\leq_s$ because for no $w \in a^*b^*c^*$, $abca \leq_s w$.

4. Consider two morphisms $g, h : \Sigma^* \longrightarrow \Delta^*$ and the associated equality set

$$E(g,h) = \{w \in \Sigma^+ \mid g(w) = h(w)\}.$$

Because $E(g,h)$ is closed under catenation, it follows that the language $E(g,h)$ is confluent w.r.t. any of the quasi orders $\leq_f, \leq_s$, and $\leq_\mathcal{P}$.

In fact, this a particular case of a more general result: if a language L is a star language, $L = L^+$, then L is confluent w.r.t. any quasi order $\leq \supseteq \leq_f$.

3 Well quasi orders

The concept of a well quasi order has been frequently discovered; see, for instance, [ER], [Hi], [Kr1], [Ha]. A complete account in this matter is given by [Kr2]. The definitions we use here are from [CK].

Let $\leq$ be a quasi order on Σ^*. A set $L \subseteq \Sigma^*$ is an *antichain* of $\leq$ if all elements in L are pairwise incomparable w.r.t. $\leq$, that is, for any $u, v \in L$, neither $u \leq v$ nor $v \leq u$. The quasi order $\leq$ is called *well founded* if there is no infinite descending sequence $w_1 \leq^{-1} w_2 \leq^{-1} w_3 \leq^{-1} \cdots$ such that, for no $i \geq 1, w_i \leq w_{i+1}$, where $\leq^{-1}$ denotes the inverse of $\leq$. If $\leq$ is well founded and any antichain of it is finite, then it is a *well quasi order*. The quasi order $\leq$ is *monotone* if, for any $x, y, u, v \in \Sigma^*$, $x \leq y$ implies $uxv \leq uyv$.

We mention some results on well quasi orders which we shall need later on. Some of them can be also found in [dLV] and [Lo].

The first one is a particular form of the well-known theorem by Higman, [Hi], which easily implies the fact the the subword partial order is a well partial order.

Theorem 3.1 (Higman, [Hi]) *If L is a language such that any two words in L are incomparable w.r.t. the subword partial order $\leq_s$, then L is finite.*

Corollary 3.2 *The Parikh subword quasi order $\leq_\mathcal{P}$ is a well quasi order.*

Haines, [Ha], seems to have been the first to notice that both sets, of subwords and of superwords, are regular for any language.

Theorem 3.3 (Haines, [Ha]) *For any language L, both sets $\downarrow_{\leq_s} L$ and $\downarrow_{\leq_s^{-1}} L$ are regular.*

As shown in [HI], this holds for any monotone well quasi order, in particular for the Parikh subword quasi order.

Theorem 3.4 ([HI]) *For any monotone well quasi order $\leq$ on Σ^* and language $L \subseteq \Sigma^*$, the sets $\downarrow_{\leq} L$ and $\downarrow_{\leq^{-1}} L$ are regular.*

Corollary 3.5 *For any language L, the sets $\downarrow_{\leq_{\mathcal{P}}} L$ and $\downarrow_{\leq_{\mathcal{P}}^{-1}} L$ are regular.*

It may happen that regular expressions for $\downarrow_{\leq} L$ and $\downarrow_{\leq^{-1}} L$ cannot be found effectively. As noticed by van Leeuwen in [vL], if this is possible for a family of languages, then the emptiness problem is decidable for the same family of languages. The converse of this statement has been shown to be true in the case of $\downarrow_{\leq_s^{-1}} L$ for families effectively closed under intersection with regular sets in [vL].

Theorem 3.6 (van Leeuwen, [vL]) *For any family $\mathcal{L}$ effectivelly closed under intersection with regular sets, one can effectively determine $\downarrow_{\leq_s^{-1}} L$ for each $L \in \mathcal{L}$ if and only if $\mathcal{L}$ has a uniformly decidable emptiness problem.*

Solving an open problem of Haines [Ha], van Leeuwen, [vL], proved also the following result which we shall need later on.

Theorem 3.7 (van Leeuwen, [vL]) *For context-free languages L, one can effectively determine $\downarrow_{\leq_s} L$.*

For other formal language theory notions and results we refer to [HU] and [Sa1].

4 Subwords of infinite words

We deal in this section with languages consisting of subwords of infinite words. We prove that it is decidable whether or not a given regular language is the set of subwords of a right-, left-, or bi-infinite word. The same problem for context-free languages is undecidable.

We make the following conventions which will be used in sections 4 and 5:

- whenever we write $L \subseteq \Sigma^*$, Σ is supposed to be the minimal alphabet with this property,

- $\max_{w \in \emptyset} |w| = 0$.

Also, for a finite word $w \in \Sigma^*$, we denote by $Sub(w)$ the set of subwords of w, that is, $Sub(w) = \downarrow_{\leq_s} \{w\}$.

We consider first right-infinite words. The next lemma is a characterization for a language to be the set of subwords of a right-infinite word.

Lemma 4.1 *For any language $L \subseteq \Sigma^*$, $L \in \mathcal{F}^r_{\leq_s}$ if and only if*

– either $L = \Sigma^*$ and, for any $a \in \Sigma$, $card(\Sigma^* a \cap L) = \infty$,

– or else $L = Sub(w)\Sigma'^*$, for some $w \in \Sigma^M$, $\emptyset \neq \Sigma' \subset \Sigma$, and there is an $a \in \Sigma$ such that $card(\Sigma^* a \cap L) < \infty$, where

$$\Sigma' = \{a \in \Sigma \mid card(\Sigma^* a \cap L) = \infty\},$$
$$M = \max_{a \in \Sigma \setminus \Sigma'} \max_{\substack{ua \in L \\ u \in \Sigma^*}} |ua|.$$

Proof. Suppose that $L = \downarrow_{\leq_s} Pref(\alpha)$, for some $\alpha \in \Sigma^\omega$. Denote $\alpha = \alpha_1\alpha_2\alpha_3\ldots$, $\alpha_i \in \Sigma$, for any $i \geq 1$.

Suppose first that $\Sigma' = \Sigma$ and let us prove that $L = \Sigma^*$. Take $w = a_1a_2\ldots a_n \in \Sigma^*$. If $n = 0$, then $w = \lambda \in \downarrow_{\leq_s} Pref(\alpha) = L$. Assume that $n \geq 1$. Since $a_1 \in \Sigma'$, we have that card$(\Sigma^* a_1 \cap L) = \infty$ and so there is $i_1 \geq 1$ such that $\alpha_{i_1} = a_1$. Now, $a_2 \in \Sigma'$, so card$(\Sigma^* a_2 \cap L) = \infty$ and we can find $i_2 > i_1$ with $\alpha_{i_2} = a_2$. Continuing inductively, if $i_1 < i_2 < \ldots < i_k$ were found, for some $2 \leq k \leq n-1$, then, again because card$(\Sigma^* a_{k+1} \cap L) = \infty$, we find $i_{k+1} > i_k$ such that $\alpha_{i_{k+1}} = a_{k+1}$. Thus $w = \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_n} \in \downarrow_{\leq_s} Pref(\alpha) = L$.

Suppose now that $\Sigma' \subset \Sigma$. With the notations above, we claim that

$$L = Sub(\alpha_1\alpha_2\ldots\alpha_M)\Sigma'^*.$$

We will establish inclusion in both directions.

Notice that, by the definition of $M, \alpha_n \in \Sigma'$, for any $n > M$. Take $w \in L, w \neq \lambda$ and put $w = \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_n}, n \geq 1, 1 \leq i_1 < i_2 < \ldots < i_n$. If $i_n \leq M$, then $w \in Sub(\alpha_1\alpha_2\ldots\alpha_M)$. If $M < i_1$, then $\alpha_{i_j} \in \Sigma'$, for any $1 \leq j \leq n$, and so $w \in \Sigma'^*$. If, for some $1 \leq k < n$, $i_k \leq M < i_{k+1}$, then $\alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_k} \in Sub(\alpha_1\alpha_2\ldots\alpha_M)$ and $\alpha_{i_{k+1}}\alpha_{i_{k+2}}\ldots\alpha_{i_n} \in \Sigma'^*$. One inclusion is proved.

For the converse inclusion, take $w = \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_n} \in Sub(\alpha_1\alpha_2\ldots\alpha_M)\Sigma'^*$. We can drop an eventual prefix of w in $Sub(\alpha_1\alpha_2\ldots\alpha_M)$ and it is enough to show that $\Sigma'^* \subseteq \downarrow_{\leq_s} Pref(\alpha_{M+1}\alpha_{M+2}\ldots)$. This is done using the same inductive argument as above.

For the converse implication in our statement, let us put $\Sigma = \{a_1, a_2, \ldots, a_n\}$. If $L = \Sigma^*$, then

$$L = \downarrow_{\leq_s} Pref((a_1a_2\ldots a_n)^\omega),$$

so $L \in \mathcal{F}^r_{\leq_s}$. If $L = Sub(w)\Sigma'^*$, then, if $\Sigma' = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\} \subset \Sigma$ (remember that $\Sigma' \neq \emptyset$), then

$$L = \downarrow_{\leq_s} Pref(w(a_{i_1}a_{i_2}\ldots a_{i_k})^\omega),$$

so again $L \in \mathcal{F}^r_{\leq_s}$.

□

From Lemma 4.1 we get the following corollary.

Corollary 4.2 *It is decidable whether or not an arbitrary regular language is in the family* $\mathcal{F}^r_{\leq_s}$.

Proof. Consider a regular language $R \subseteq \Sigma^*$. If R is finite, then $R \notin \mathcal{F}^r_{\leq_s}$, so suppose R infinite. Using the notations in Lemma 4.1, we can effectively determine Σ'. Now, if $\Sigma = \Sigma'$, then $R \in \mathcal{F}^r_{\leq_s}$ if and only if $R = \Sigma^*$ which is decidable. If $\Sigma' \subset \Sigma$, then $R \in \mathcal{F}^r_{\leq_s}$ if and only if $R = Sub(w)\Sigma'^*$, for some $w \in \Sigma^M$. Since M is effectively computable and there are only finitely many words in Σ^M, this is also decidable.

□

By left-right duality, we get the next result.

Corollary 4.3 *It is decidable whether or not an arbitrary regular language is in the family* $\mathcal{F}^l_{\leq_s}$.

We now turn our attention to the bi-infinite case and give a characterization for a language to be the set of subwords of a bi-infinite word.

Lemma 4.4 *A language* $L \subseteq \Sigma^*$ *belongs to the family* $\mathcal{F}^{bi}_{\leq_s}$ *if and only if*

– *either* $L = \Sigma^*$ *and, for any* $a, b \in \Sigma$, $card(a\Sigma^* b \cap L) = \infty$,

– *or* $L = \Sigma_1^*\Sigma_2^*$, *for some* $\Sigma_1 \subset \Sigma, \Sigma_2 \subset \Sigma$, *and there are* $a, b \in \Sigma, a \neq b$, *such that* $card(a\Sigma^* b \cap L) < \infty$, *for any such* a, b, $a\Sigma^* b \cap L = \emptyset$, *and, moreover, for any* $a \in \Sigma$, $card(a\Sigma^* a \cap L) = \infty$.

– *or else* $L = \Sigma_1^* Sub(w)\Sigma_2^*$, *for some* $\emptyset \neq \Sigma_1 \subset \Sigma$, $\emptyset \neq \Sigma_2 \subset \Sigma, w \in \Sigma^M$, *and there are* $a, b \in \Sigma, a \neq b$, *such that* $0 < card(a\Sigma^* b \cap L) < \infty$ *or there is an* $a \in \Sigma$ *such that* $card(a\Sigma^* a \cap L) < \infty$, *where*

$$M = \max\Big(\max_{\substack{a,b\in\Sigma\\ card(a\Sigma^* b\cap L)<\infty}} \max_{\substack{aub\in L\\ u\in\Sigma^*}} |aub|, \max_{\substack{a\in\Sigma\\ card(a\Sigma^* a\cap L)<\infty}} \big(\max_{\substack{aua\in L\\ u\in\Sigma^*}} |aua|, |a|\big)\Big).$$

(Observe that $|a|$ *is needed in the definition of* M *for the case when there is an* $a \in \Sigma$ *such that* $card(b\Sigma^* c \cap L) = \infty$, *whenever at least one of* b, c *is different from* a, *and there is exactly one occurrence of* a *in any bi-infinite word* α *with* $L = \downarrow_{\leq_s} Fact(\alpha)$.*)*

Proof. It can be shown as in the proof of Lemma 4.1 that if one of the three conditions in the statement is fulfilled, then a word $\alpha \in {}^\omega\Sigma^\omega$ such that $L = \downarrow_{\leq_s} Fact(\alpha)$ can be constructed.

Let us prove the converse implication. Suppose that $L \in \mathcal{F}^{bi}_{\leq_s}$ and put $L = \downarrow_{\leq_s} Fact(\alpha)$, for some $\alpha \in {}^\omega\Sigma^\omega$. Denote $\alpha = \ldots\alpha_{-2}\alpha_{-1}\alpha_0\alpha_1\alpha_2\ldots$. (The notation is not unique, but we consider an arbitrary fixed one.) Suppose first that, for any $a, b \in \Sigma$, $\mathrm{card}(a\Sigma^* b \cap L) = \infty$ and let us prove that $L = \Sigma^*$.

In particular, for any $a \in \Sigma$, $\mathrm{card}(a\Sigma^* a \cap L) = \infty$, thus either $m_+(a) = \sup_{\substack{n \in \mathbf{Z}_+ \\ \alpha_n = a}} n = \infty$, or $m_-(a) = \sup_{\substack{n \in \mathbf{Z}_- \\ \alpha_n = a}} (-n) = \infty$, or both. Moreover, if, for some $a \in \Sigma$, one of the two suprema above is finite, then for any $b \in \Sigma$ the other supremum corresponding to b is infinite. Indeed, if for instance $m_-(a) < \infty$, then from $\mathrm{card}(a\Sigma^* b \cap L) = \infty$ we get $m_+(b) = \infty$.

Take now $a_1 a_2 \ldots a_n \in \Sigma^*$ an arbitrary non-empty word. We have two cases: either $m_+(a_i) = m_-(a_i) = \infty$, for every $1 \le i \le n$, or, for some $i_0, 1 \le i_0 \le n$, one of $m_+(a_{i_0})$ and $m_-(a_{i_0})$ is finite. Since the reasoning in the second case applies also for the first one, we consider the second case only. We may suppose, without loss of generality, that $m_-(a_{i_0}) < \infty$. It follows that $m_+(a_i) = \infty$, for any $1 \le i \le n$. Now, by our usual inductive procedure, we find the integers $k_1 < k_2 < \ldots < k_n$ such that $a_i = \alpha_{k_i}, 1 \le i \le n$. Thus $a_1 a_2 \ldots a_n \in \downarrow_{\le_s} Fact(\alpha) = L$.

Suppose next that there are $a, b \in \Sigma, a \ne b$, such that $\mathrm{card}(a\Sigma^* b \cap L) < \infty$ and for any such a and b, $\mathrm{card}(a\Sigma^* b \cap L) = 0$. Choose some fixed $a, b \in \Sigma, a \ne b$, such that $\mathrm{card}(a\Sigma^* b \cap L) = 0$. This means that there are no $m, n \in \mathbf{Z}, m < n$, such that $\alpha_m = a$ and $\alpha_n = b$. Suppose also that $\mathrm{card}(c\Sigma^* c \cap L) = \infty$, for any $c \in \Sigma$.

We construct the sets Σ_1 and Σ_2 as follows:

$$\Sigma_1 = \{c \in \Sigma \mid m_-(c) = \infty\},$$
$$\Sigma_2 = \{c \in \Sigma \mid m_+(c) = \infty\}.$$

Notice that $a \notin \Sigma_1$ and $b \notin \Sigma_2$, so $\Sigma_1 \subset \Sigma, \Sigma_2 \subset \Sigma$.

We claim that $L = \Sigma_1^* \Sigma_2^*$. From the construction of Σ_1 and Σ_2, it follows easily that $\Sigma_1^* \Sigma_2^* \subseteq L$.

For the converse inclusion, take $a_1 a_2 \ldots a_n \in L$. Then, for any $1 \le i \le n-1$, $\mathrm{card}(a_i \Sigma^* a_{i+1} \cap L) = \infty$, thus either $m_-(a_i) = \infty$ or $m_+(a_{i+1}) = \infty$. Therefore, we have three cases:

(i) For any $1 \le i \le n, m_-(a_i) = \infty$. Then, $a_i \in \Sigma_1$, for any $1 \le i \le n$, hence $a_1 a_2 \ldots a_n \in \Sigma_1^* \subseteq \Sigma_1^* \Sigma_2^*$.

(ii) For any $1 \le i \le n, m_+(a_i) = \infty$, which is similar to (i).

(iii) There is $1 \le i_0 \le n-1$ such that $m_-(a_i) = \infty$, for any $1 \le i \le i_0$, and $m_+(a_i) = \infty$, for any $i_0 + 1 \le i \le n$. Then $a_1 a_2 \ldots a_{i_0} \in \Sigma_1^*$ and $a_{i_0+1} a_{i_0+2} \ldots a_n \in \Sigma_2^*$, hence $a_1 a_2 \ldots a_n \in \Sigma_1^* \Sigma_2^*$. The equality is proved.

Finally, suppose that either there are $a, b \in \Sigma, a \ne b$, with $0 < \mathrm{card}(a\Sigma^* b \cap L) < \infty$ or, for some $a \in \Sigma, \mathrm{card}(a\Sigma^* a \cap L) < \infty$. Notice that, due to the minimality of Σ, for any $a \in \Sigma, \mathrm{card}((a\Sigma^* a \cup \{a\}) \cap L) > 0$.

Suppose first that $M \ge 2$ and the maximum

$$\max\Big(\max_{\substack{a,b \in \Sigma \\ \mathrm{card}(a\Sigma^* b \cap L) < \infty}} \max_{\substack{aub \in L \\ u \in \Sigma^*}} |aub|, \max_{\substack{a \in \Sigma \\ \mathrm{card}(a\Sigma^* a \cap L) < \infty}} \max_{\substack{aua \in L \\ u \in \Sigma^*}} |aua| \Big)$$

is reached for the word $w = cvd \in L, |w| = M$. Then, for some $n_0 \in \mathbf{Z}, \alpha_{n_0} = c, \alpha_{n_0+1}\alpha_{n_0+2}\ldots\alpha_{n_0+M-2} = v, \alpha_{n_0+M-1} = d$. Put

$$\Sigma_1 = \{a \in \Sigma \mid \text{ there is } n \leq n_0 - 1 \text{ such that } \alpha_n = a\},$$
$$\Sigma_2 = \{a \in \Sigma \mid \text{ there is } n \geq n_0 + M \text{ such that } \alpha_n = a\}.$$

We claim that $L = \Sigma_1^* Sub(w)\Sigma_2^*$. The inclusion $L \subseteq \Sigma_1^* Sub(w)\Sigma_2^*$ is clear from the definition of Σ_1 and Σ_2. For the converse inclusion, it is enough to see that, for any $a \in \Sigma_1$, $m_-(a) = \infty$ and, for any $a \in \Sigma_2$, $m_+(a) = \infty$.

For the case $M = 1$, there is a unique $a \in \Sigma$ which occurs exactly once in α (see also the remark at the end of the statement). By an eventual translation, we may suppose that $\alpha_0 = a$. Now, for

$$\Sigma_1 = \{b \in \Sigma \mid \text{ there is } n < 0 \text{ such that } \alpha_n = b\},$$
$$\Sigma_2 = \{b \in \Sigma \mid \text{ there is } n > 0 \text{ such that } \alpha_n = b\},$$

we have as above $L = \Sigma_1^* Sub(a)\Sigma_2^*$. The proof is concluded.

□

From Lemma 4.4 we get

Corollary 4.5 *It is decidable whether or not an arbitrary regular language is in the family* $\mathcal{F}^{bi}_{\leq_s}$.

Proof. All conditions in Lemma 4.4 are decidable for regular languages.

□

From Corollaries 4.2, 4.3, and 4.5, we obtain the main result of this section.

Theorem 4.6 *It is decidable whether or not an arbitrary regular language is in the family* $\mathcal{F}_{\leq_s}$.

The same problem is undecidable for context-free languages.

Theorem 4.7 *It is undecidable whether or not an arbitrary context-free language is in the family* $\mathcal{F}_{\leq_s}$.

Proof. Consider an arbitrary Turing machine M and its set of invalid computations L_M. It is known (see, for instance, [HU]) that L_M is context-free. We claim that L_M belongs to the family $\mathcal{F}_{\leq_s}$ if and only if its complement is empty.

One implication is obvious. For the other one, suppose that $L_M \in \mathcal{F}_{\leq_s}$. It follows that any subword of a word in L_M is also in L_M. Thus, any subword of an invalid computation is also an invalid computation. Since any valid computation is a subword of an invalid one, it follows that there is actually no valid computation, hence $\overline{L_M} = \emptyset$.

Consequently, $L_M \in \mathcal{F}_{\leq_s}$ if and only if the set of valid computations of M is empty which is equivalent to the fact that the language accepted by M is empty which is undecidable. The result follows.

□

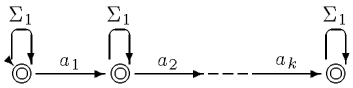
Observe that, rather than containing new aspects about the family $\mathcal{F}_{\leq_s}$, Theorem 4.7 is another manifestation of the undecidability of the equality between a given context-free language and Σ^*. Theorem 4.7, as opposed to Theorem 4.6, shows the power of the devices defining context-free languages, not the power of the languages themselves.

5 Parikh subwords

We consider in this section the Parikh subword quasi order. We prove that the confluence problem w.r.t. $\leq_{\mathcal{P}}$ is decidable for regular languages. From this, we deduce that it is also decidable whether or not an arbitrary regular language is in the family $\mathcal{F}_{\leq_{\mathcal{P}}}$. Using the effective regularity of down-sets w.r.t. $\leq_{\mathcal{P}}$ for context-free languages, we prove that the confluence problem w.r.t. $\leq_{\mathcal{P}}$ is decidable even in the context-free case.

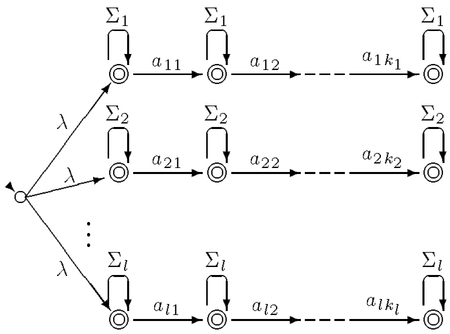
It is known, see Corollary 3.5, that, for any language R, the down-set $\downarrow_{\leq_{\mathcal{P}}} R$ is regular. We show that $\downarrow_{\leq_{\mathcal{P}}} R$ is effectively regular and a finite automaton of a special form can be constructed for it. We need some definitions.

A *chain automaton* over the alphabet Σ is a finite nondeterministic (incomplete) automaton defined by the graph



where $k \geq 0, a_i \in \Sigma, 1 \leq i \leq k, \Sigma_1 \subseteq \Sigma$, and $\Sigma_1 \cap \{a_1, a_2, \ldots, a_k\} = \emptyset$. Observe that all states are final and Σ_1 may be empty.

A *multichain automaton* is defined similarly:



The notation is the same as before: each Σ_i is a subset of Σ and each a_{ij} is a letter of Σ. Also $\Sigma_i \cap \{a_{i1}, a_{i2}, \ldots, a_{ik_i}\} = \emptyset$, for any $1 \leq i \leq l$, and all states, except the initial state, are final.

A multichain automaton is *chain-reduced* if none of its chains is superfluous in the sense that every word accepted by the chain is accepted by some of the other chains as well.

The following lemma is obvious, since inclusion is decidable for regular languages.

Lemma 5.1 *For a given multichain automaton, an equivalent chain-reduced multichain automaton can be effectively constructed.*

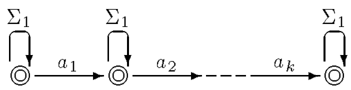
Lemma 5.2 *For a given regular language R, a chain-reduced multichain automaton accepting the language $\downarrow_{\leq_{\mathcal{P}}} R$ can be effectively constructed.*

Proof. The proof is by induction on the regular expression defining R. It suffices to construct a multichain automaton for $\downarrow_{\leq_{\mathcal{P}}} R$; the chain-reduction can be carried out using the preceding lemma.

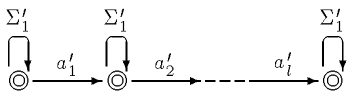
The basis of the induction is clear: the statement holds if R is a letter or $\emptyset$. (Observe that $\lambda = \emptyset^*$.) We make the inductive hypothesis: the statement holds true if R is defined by a regular expression α or by a regular expression β. We show that it holds true if R is defined by any of $(\alpha + \beta), (\alpha\beta)$ or α^*.

As regards union, we only have to combine the two multichain automata in an obvious fashion. (Recall that at this stage we do not have to worry about the result being chain-reduced.)

Consider $(\alpha\beta)$. The multichain automaton for it is constructed as follows. Let



be a chain in the automaton for α and



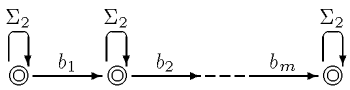
a chain in the automaton for β. Put

$$\Sigma_2 = \Sigma_1 \cup \Sigma_1'$$

and

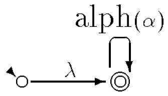
$$\{b_1, b_2, \ldots, b_m\} = (\{a_1, a_2, \ldots, a_k\} \cup \{a_1', a_2', \ldots, a_l'\}) \backslash \Sigma_2.$$

Then



as well as every permutation of it, is a chain in the new automaton.

Finally, consider α^*. Let $\mathrm{alph}(\alpha) \subseteq \Sigma$ be the set of letters appearing in α. The multichain automaton for α^* is simply



□

Remark. The loop complexity of a regular language may reduce drastically in the construction of Lemma 5.2. For instance, each of the languages in the sequence considered in [DS] or [Sa2], is reduced to $\{a, b\}^*$, although the star height of the languages in the sequence is not bounded.

From Lemma 5.2 and Theorem 3.7, we deduce the following corollary.

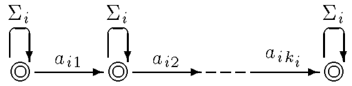
Corollary 5.3 *For any context-free language L, the down-set $\downarrow_{\leq_\mathcal{P}} L$ is effectively regular.*

It has been proved in [HI] that the confluence property w.r.t. any quasi order translates upwards, from a language to its down-set, and downwards, conversely.

Theorem 5.4 ([HI]) *For any quasi order $\leq$, a language L is confluent w.r.t. $\leq$ if and only if the down-set $\downarrow_{\leq} L$ is confluent w.r.t. $\leq$.*

Theorem 5.5 *It is decidable whether or not an arbitrary regular language is confluent w.r.t. $\leq_\mathcal{P}$.*

Proof. Let R be a regular language. Using Lemma 5.2, we construct a chain-reduced multichain automaton accepting the language $\downarrow_{\leq_{\mathcal{P}}} R$. Suppose that this automaton consists of n chains, $c_1, c_2, \ldots, c_n$, and, for any $1 \leq i \leq n, c_i$ is defined by the graph



where $\Sigma_i = \{b_{i1}, b_{i2}, \ldots, b_{il_i}\}$.

Notice that each c_i is uniquely identified by the word $w_i = a_{i1}a_{i2}\ldots a_{ik_i}$ and the set Σ_i.

Claim. $\downarrow_{\leq_{\mathcal{P}}} R$ is confluent w.r.t. $\leq_{\mathcal{P}}$ if and only if, for any $1 \leq i, j \leq n$, $\Sigma_i = \Sigma_j$ and the words w_i and w_j are a permutation of each other.

Proof of Claim. If the conditions on c_i's in the statement of our claim are fulfilled, then clearly $\downarrow_{\leq_{\mathcal{P}}} R$ is confluent w.r.t. $\leq_{\mathcal{P}}$.

Conversely, suppose that $\downarrow_{\leq_{\mathcal{P}}} R$ is confluent w.r.t. $\leq_{\mathcal{P}}$ and consider two arbitrary $i, j, 1 \leq i, j \leq n$. Denote

$$M = \max_{1 \leq i \leq n} |w_i| + 1$$

and consider the words

$$\begin{aligned} u_i &= w_i b_{i1}^M b_{i2}^M \ldots b_{il_i}^M, \\ u_j &= w_j b_{j1}^M b_{j2}^M \ldots b_{jl_j}^M. \end{aligned}$$

Since $\downarrow_{\leq_{\mathcal{P}}} R$ is confluent w.r.t. $\leq_{\mathcal{P}}$, there is $w \in \downarrow_{\leq_{\mathcal{P}}} R$ such that $u_i \leq_{\mathcal{P}} w, u_j \leq_{\mathcal{P}} w$. Suppose that w is recognized by the chain c_m, for some $1 \leq m \leq n$. From the choice of M it follows that any word $b_{ir}^M, 1 \leq r \leq l_i$, must be generated, at least partially, in a loop. Thus $\Sigma_i \subseteq \Sigma_m$ and we claim that $\Sigma_i = \Sigma_m$. If this is not the case, then take $b \in \Sigma_m \backslash \Sigma_i$. There is a permutation of w, say w', such that $u_i \leq_s w'$. Since the word $w'b^M$ belongs to $\downarrow_{\leq_{\mathcal{P}}} R$, it is recognized by a chain different from c_i, say c_r. It follows that any word recognized by c_i is recognized by c_r as well, a contradiction. Therefore $\Sigma_i = \Sigma_m$.

Now $w_i \leq_{\mathcal{P}} w_m$ and it can be proved by contradiction as above that w_i and w_m must be a permutation of each other.

The same reasoning can be done with k instead of i. Thus, we have $\Sigma_i = \Sigma_j$ and w_i is a permutation of w_j. The claim is proved.

Since the conditions on c_i's in our claim above can be effectively checked, it follows that it can be decided whether or not $\downarrow_{\leq_{\mathcal{P}}} R$ is confluent w.r.t. $\leq_{\mathcal{P}}$. As, by Theorem 5.4, R and $\downarrow_{\leq_{\mathcal{P}}} R$ are simultaneously confluent w.r.t. $\leq_{\mathcal{P}}$, the proof is concluded.

□

Notice that it can be proved as in Theorem 5.5 that it is decidable whether or not an arbitrary regular language belongs to the family $\mathcal{F}_{\leq_{\mathcal{P}}}$ but we will do this using the following lemma which establishes a connection between the property of a language being confluent w.r.t. $\leq_{\mathcal{P}}$ and belonging to the family $\mathcal{F}_{\leq_{\mathcal{P}}}$.

Lemma 5.6 *A language $L \subseteq \Sigma^*$ is in the family $\mathcal{F}_{\leq_{\mathcal{P}}}$ if and only if the following conditions are fulfilled:*
(i) L is infinite;
(ii) $L = \downarrow_{\leq_{\mathcal{P}}} L$;
(iii) L is confluent w.r.t. $\leq_{\mathcal{P}}$.

Proof. If $L \in \mathcal{F}_{\leq_{\mathcal{P}}}$, then it is easy to see that L is infinite and $L = \downarrow_{\leq_{\mathcal{P}}} L$. To show that L is confluent w.r.t. $\leq_{\mathcal{P}}$, put $L = \downarrow_{\leq_{\mathcal{P}}} Fact(\alpha)$, for some $\alpha \in {}^{\omega}\Sigma \cup \Sigma^{\omega} \cup {}^{\omega}\Sigma^{\omega}$ and take $u, v \in L$. Then $u = \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_r}$, $v = \alpha_{j_1}\alpha_{j_2}\ldots\alpha_{j_s}$, $\alpha_{i_k}, \alpha_{j_l} \in \Sigma$, and, if $w = \alpha_{k_1}\alpha_{k_2}\ldots\alpha_{k_t}$, $\{k_1, \ldots, k_t\} = \{i_1, \ldots, i_r\} \cup \{j_1, \ldots, j_s\}$, then $w \in L, u \leq_{\mathcal{P}} w$, and $v \leq_{\mathcal{P}} w$.

Let us prove the converse implication. Suppose that $L \subseteq \Sigma^* = \{a_1, a_2, \ldots a_n\}^*$ with the three properties (i)–(iii) in the statement.

Consider the projections $\pi_i : \Sigma^* \longrightarrow \{a_i\}^*, 1 \leq i \leq n$, given by

$$\pi_i(a_j) = \begin{cases} a_i, & \text{if } i = j, \\ \lambda, & \text{if } i \neq j. \end{cases}$$

We put

$$\begin{aligned} \Sigma_1 &= \{a_i \in \Sigma \mid \pi_i(L) \text{ is infinite }\} = \{a_{j_1}, a_{j_2}, \ldots, a_{j_k}\}, \\ \Sigma_2 &= \Sigma \backslash \Sigma_1 = \{a_{i_1}, a_{i_2}, \ldots, a_{i_l}\}, \end{aligned}$$

for some $k \geq 1, l \geq 0$, and $1 \leq j_r \leq n, 1 \leq i_s \leq n$, for any $1 \leq r \leq k, 1 \leq s \leq l$. (Notice that, because L is infinite, Σ_1 is non-empty.)

Consider also

$$\max_{w \in \pi_{i_s}(L)} |w| = m_s < \infty, \text{ for any } 1 \leq s \leq l.$$

We claim that

$$L = \downarrow_{\leq_{\mathcal{P}}} Pref(a_{i_1}^{m_1} a_{i_2}^{m_2} \ldots a_{i_l}^{m_l} (a_{j_1} a_{j_2} \ldots a_{j_k})^{\omega}).$$

The part "$\subseteq$" is clear since no word in L can contain more than m_s occurrences of the symbol a_{i_s}, for any $1 \leq s \leq l$.

In order to prove the converse inclusion, "$\supseteq$", consider a word w in the right hand member of the equality to be proved. Now, for any $1 \leq s \leq l$, there is a word $w_{i_s} \in L$ with $|w_{i_s}|_{a_{i_s}} \geq |w|_{a_{i_s}}$, since there are in L words containing m_s occurrences of a_{i_s} and w cannot contain more. Also, for any $1 \leq r \leq k$, there is a word $w_{j_r} \in L$ with $|w_{j_r}|_{a_{j_r}} \geq |w|_{a_{j_r}}$, because there are in L words containing arbitrarily many occurrences of a_{j_r}. Since L is confluent w.r.t. $\leq_{\mathcal{P}}$, we get inductively a word $w' \in L$ such that

$$\begin{array}{l} w_{i_s} \leq_{\mathcal{P}} w', \quad 1 \leq s \leq l, \\ w_{j_r} \leq_{\mathcal{P}} w', \quad 1 \leq r \leq k, \end{array}$$

and so $w \leq_{\mathcal{P}} w'$. Thus $w \in \downarrow_{\leq_{\mathcal{P}}} L = L$ and the equality is proved.

□

Notice that a similar result does not hold for the subword partial order $\leq_s$. Indeed, the language $L = a^*b^*c^*$ is infinite, $L = \downarrow_{\leq_s} L$, and it is confluent w.r.t. $\leq_s$. But, clearly, L is not the set of subwords of a right-, left-, or bi-infinite word.

From the previous results in this section, we obtain

Theorem 5.7 *It is decidable whether or not an arbitrary regular language belongs to the family $\mathcal{F}_{\leq_{\mathcal{P}}}$.*

Proof. Let R be a regular language. If R is finite, then it does not belong to the family $\mathcal{F}_{\leq_{\mathcal{P}}}$. Suppose then that R is infinite. By Lemma 5.6, if $R \neq \downarrow_{\leq_{\mathcal{P}}} R$, then $R \notin \mathcal{F}_{\leq_{\mathcal{P}}}$. Since, using Lemma 5.2, it can be decided whether or not $R = \downarrow_{\leq_{\mathcal{P}}} R$, we can restrict the problem to languages R with $R = \downarrow_{\leq_{\mathcal{P}}} R$. Now, by Lemma 5.6, $R \in \mathcal{F}_{\leq_{\mathcal{P}}}$ if and only if R is confluent w.r.t. $\leq_{\mathcal{P}}$, which is decidable by Theorem 5.5.

□

From the proof of Lemma 5.6 we get another consequence.

Corollary 5.8 $\mathcal{F}^{l}_{\leq_{\mathcal{P}}} = \mathcal{F}^{r}_{\leq_{\mathcal{P}}} = \mathcal{F}^{bi}_{\leq_{\mathcal{P}}} = \mathcal{F}_{\leq_{\mathcal{P}}}$.

Proof. We show that, for any language $L \subseteq \Sigma^*, L$ belongs to any of the four families in the statement if and only if the conditions (i)-(iii) in Lemma 5.6 are fulfilled. As seen in the proof there, if L belongs to any of the families, then (i)-(iii) are fulfilled. Conversely, using the notations in the proof of Lemma 5.6, we have

$$\begin{aligned} L &= \downarrow_{\leq_{\mathcal{P}}} Pref(a_{i_1}^{m_1} a_{i_2}^{m_2} \ldots a_{i_l}^{m_l} (a_{j_1} a_{j_2} \ldots a_{j_k})^{\omega}) = \\ &= \downarrow_{\leq_{\mathcal{P}}} Suf({}^{\omega}(a_{j_1} a_{j_2} \ldots a_{j_k}) a_{i_1}^{m_1} a_{i_2}^{m_2} \ldots a_{i_l}^{m_l}) = \\ &= \downarrow_{\leq_{\mathcal{P}}} Fact({}^{\omega}(a_{j_1} a_{j_2} \ldots a_{j_k}) a_{i_1}^{m_1} a_{i_2}^{m_2} \ldots a_{i_l}^{m_l} (a_{j_1} a_{j_2} \ldots a_{j_k})^{\omega}). \end{aligned}$$

The equalities in the statement are now clear.

$\square$

The last result is particular for Parikh subwords. It does not hold for any quasi order $\leq \subseteq \leq_s$, a fact not surprising because of the permutation of letters allowed in the Parikh subword quasi order. Indeed, we have the following examples: $a^*b^* \in \mathcal{F}^{bi}_{\leq} \backslash (\mathcal{F}^{l}_{\leq} \cup \mathcal{F}^{r}_{\leq})$, $a^*b \in \mathcal{F}^{l}_{\leq} \backslash (\mathcal{F}^{r}_{\leq} \cup \mathcal{F}^{bi}_{\leq})$, and $ba^* \in \mathcal{F}^{r}_{\leq} \backslash (\mathcal{F}^{l}_{\leq} \cup \mathcal{F}^{bi}_{\leq})$.

We can prove now the following rather unexpected result.

Theorem 5.9 *It is decidable whether or not an arbitrary context-free language is confluent w.r.t.* $\leq_{\mathcal{P}}$.

Proof. Let L be a context-free language. By Corollary 5.3, the down-set $\downarrow_{\leq_{\mathcal{P}}} L$ is effectively regular and by Theorem 5.4, L is confluent w.r.t. $\leq_{\mathcal{P}}$ if and only if $\downarrow_{\leq_{\mathcal{P}}} L$ is confluent w.r.t. $\leq_{\mathcal{P}}$, which is decidable by Theorem 5.5.

$\square$

Notice that the next result can be proved as Theorem 4.7. Remarks similar to our comparison between Theorems 4.6 and 4.7 can be made in this case as well.

Theorem 5.10 *It is undecidable whether or not an arbitrary context-free language is in the family* $\mathcal{F}_{\leq_{\mathcal{P}}}$.

6 Generalized confluence and well quasi orders

In this section, we generalize the concept of confluence and show that the generalization is strongly connected with the concept of well quasi order.

Let $\leq$ be a quasi order on Σ^* and $k \geq 1$ an arbitrary fixed integer. A language $L \subseteq \Sigma^*$ is called *k-confluent w.r.t.* $\leq$ if and only if

$$L = \bigcup_{i=1}^{k} L_i$$

for some languages $L_i \subseteq \Sigma^*$, L_i confluent w.r.t. $\leq$, for any $1 \leq i \leq k$. (Notice that this is a generalization of the ordinary confluence since 1-confluent is the same as confluent.)

L is called *confluent w.r.t.* $\leq$ *in generalized sense* if L is k-confluent w.r.t. $\leq$, for some $k \geq 1$.

Examples

1. The language

$$L_1 = a^* \cup b^*,$$

which is not confluent w.r.t. $\leq_{\mathcal{P}}$, is 2-confluent w.r.t. $\leq_p$ since any of the two languages a^* and b^* is confluent w.r.t. $\leq_p$.

2. For any $k \geq 1$, the language

$$L_{2,k} = \{a^i b a^j \mid i, j \geq 0, i + j = k\}$$

is (k+1)-confluent w.r.t. $\leq_s$ but not k-confluent w.r.t. $\leq_s$. Notice that $L_{2,k}$ is confluent w.r.t. $\leq_{\mathcal{P}}$.

3. For any alphabet Σ with card$(\Sigma) \geq 2$, the language

$$L_3 = \Sigma^*$$

is not confluent w.r.t. $\leq_p$ in generalized sense. Indeed, consider $a, b \in \Sigma, a \neq b$. For any $k \geq 1$, any two of the words $b, ab, a^2b, \ldots, a^kb$ cannot be in the same confluent part w.r.t. $\leq_p$ of Σ^*. Therefore, Σ^* is not k-confluent w.r.t. $\leq_p$.

The next theorem states that, for any quasi order $\leq$, a language L and its down-set $\downarrow_{\leq} L$ are confluent w.r.t. $\leq$ in generalized sense at the same time. Moreover, the respective constant k is the same. This is a generalization of Theorem 5.4.

Theorem 6.1 *For any quasi order $\leq$ on Σ^* and integer $k \geq 1$, a language $L \subseteq \Sigma^*$ is k-confluent w.r.t. $\leq$ if and only if the down-set $\downarrow_{\leq} L$ is k-confluent w.r.t. $\leq$.*

Proof. Suppose first that L is k-confluent w.r.t. $\leq$. By definition, we have

$$L = \bigcup_{i=1}^{k} L_i,$$

for some languages $L_i \subseteq \Sigma^*, L_i$ confluent w.r.t. $\leq$, for any $1 \leq i \leq k$. We have

$$\downarrow_{\leq} L = \downarrow_{\leq} (\bigcup_{i=1}^{k} L_i) = \bigcup_{i=1}^{k} \downarrow_{\leq} L_i$$

and, by Theorem 5.4, any language $\downarrow_{\leq} L_i, 1 \leq i \leq k$, is confluent w.r.t. $\leq$. Thus $\downarrow_{\leq} L$ is k-confluent w.r.t. $\leq$.

Conversely, suppose that $\downarrow_{\leq} L$ is k-confluent w.r.t. $\leq$ and put

$$\downarrow_{\leq} L = \bigcup_{i=1}^{k} R_i,$$

for some languages $R_i \subseteq \Sigma^*$, R_i confluent w.r.t. $\leq$, for any $1 \leq i \leq k$. As above, we have

$$\downarrow_{\leq} L = \bigcup_{i=1}^{k} \downarrow_{\leq} R_i,$$

so we may assume that, for any $1 \leq i \leq k$, $R_i = \downarrow_{\leq} R_i$. Again, we may indeed make this assumption because, by Theorem 5.4, $\downarrow_{\leq} R_i$ is confluent w.r.t. $\leq$ since R_i is.

Suppose also that, for no $1 \leq j \leq k$,

$$\downarrow_{\leq} L = \bigcup_{\substack{i=1 \\ i \neq j}}^{k} R_i.$$

This would mean that we can eventually decrease k to $k' < k$. Then, if we prove that L is k'-confluent w.r.t. $\leq$, then it is also k-confluent w.r.t. $\leq$.

Let us define, for any $1 \leq i \leq k$, the language

$$L_i = L \cap R_i.$$

Claim 1. $L = \bigcup_{i=1}^{k} L_i$.

Proof of Claim 1. Since $L_i \subseteq L$, for any $1 \leq i \leq k$, the inclusion "$\supseteq$" follows.

For the converse inclusion, take $w \in L$. Because $\leq$ is reflexive, $L \subseteq \downarrow_{\leq} L$, so $w \in \downarrow_{\leq} L$ and thus $w \in R_i$, for some $1 \leq i \leq k$. Hence $w \in L \cap R_i = L_i$. The claim is proved.

Claim 2. For any $1 \leq i \leq k$ and $x \in R_i$, there exists $y \in L \cap R_i$ such that $x \leq y$.

Proof of Claim 2. We argue by contradiction. Suppose that there are $i_0, 1 \leq i_0 \leq k$, and $x \in R_{i_0}$ such that, for any y, $x \leq y$ implies $y \notin L \cap R_{i_0}$.

Take an arbitrary $w \in R_{i_0}$. Since R_{i_0} is confluent w.r.t. $\leq$, there is $y \in R_{i_0}$ such that $w \leq y$ and $x \leq y$. Using our assumption, it follows that $y \notin L$. But $y \in R_{i_0} \subseteq \downarrow_{\leq} L$, so $y \leq z$, for some $z \in L$. Then $z \in R_j$, for some $1 \leq j \leq k$. We must have $j \neq i_0$ because $j = i_0$ would imply $z \in L \cap R_{i_0}$ and as $x \leq z$, by the transitivity of $\leq$, this would contradict our assumption.

Consequently, $w \leq z$ and, as we supposed that $R_j = \downarrow_{\leq} R_j$, we get $w \in R_j$. Because $j \neq i_0$ and w was arbitrarily chosen in R_{i_0}, we have

$$R_{i_0} \subseteq \bigcup_{\substack{i=1 \\ i \neq i_0}}^{k} R_i$$

thus

$$\downarrow_{\leq} L = \bigcup_{\substack{i=1 \\ i \neq i_0}}^{k} R_i,$$

a contradiction which proves the claim.

Claim 3. For any $1 \leq i \leq k, \downarrow_{\leq} L_i = R_i$.

Proof of Claim 3. Since, by definition, $L_i \subseteq R_i$ and, as we supposed, $R_i = \downarrow_{\leq} R_i$, the inclusion $\downarrow_{\leq} L_i \subseteq R_i$ follows.

For the converse inclusion, take $x \in R_i$. By Claim 2, there is $y \in L \cap R_i = L_i$ such that $x \leq y$. Consequently, $x \in \downarrow_{\leq} L_i$ and the equality is proved.

Finally, the fact that R_i is confluent w.r.t. $\leq$ implies, by Claim 3, that $\downarrow_{\leq} L_i$ is confluent w.r.t. $\leq$. By Theorem 5.4, it follows that also L_i is confluent w.r.t. $\leq$. Using Claim 1, we have that L is k-confluent w.r.t. $\leq$ as claimed.

□

We now move to the interconnection between generalized confluence and well quasi orders. In one direction we have the following result.

Theorem 6.2 *If $\leq$ is a quasi order on Σ^* such that any language $L \subseteq \Sigma^*$ is confluent w.r.t. $\leq$ in generalized sense, then any antichain of $\leq$ is finite.*

Proof. Suppose that $L \subseteq \Sigma^*$ is an antichain of $\leq$. By hypothesis, L is confluent w.r.t. $\leq$ in generalized sense. Therefore, we have

$$L = \bigcup_{i=1}^{k} L_i,$$

for some $k \geq 1, L_i \subseteq \Sigma^*, L_i$ confluent w.r.t. $\leq$, for any $1 \leq i \leq k$.

Consider an $i, 1 \leq i \leq k$, and two elements $u, v \in L_i$. Since L_i is confluent w.r.t. $\leq$, there is $w \in L_i$ such that $u \leq w$ and $v \leq w$. But $u, v, w \in L$ which is an antichain of $\leq$. It follows that $u = w$ and $v = w$, thus $u = v$. As u and v were arbitrarily chosen in L_i, we get that

$$\mathrm{card}(L_i) = 1, \text{ for any } 1 \leq i \leq k.$$

Consequently,

$$\mathrm{card}(L) = k,$$

hence L is finite.

□

Corollary 6.3 *If $\leq$ is a well founded quasi order on Σ^* such that any language $L \subseteq \Sigma^*$ is confluent w.r.t. $\leq$ in generalized sense, then $\leq$ is a well quasi order.*

There is also an obvious connection between the ordinary confluence property and total orders.

Theorem 6.4 *For a partial order $\leq$ on Σ^*, $\leq$ is total if and only if any language $L \subseteq \Sigma^*$ is confluent w.r.t. $\leq$.*

In order to find a connection in the other direction, that is, to find some conditions according to which the property of being well for a quasi order $\leq$ implies the generalized confluence w.r.t. $\leq$ for any language, we need a result concerning the form of the down-sets w.r.t. $\leq_s$. It is known by Theorem 3.3 that, for any language L, the down-set $\downarrow_{\leq_s} L$ is regular.

Lemma 6.5 *For any language $L \subseteq \Sigma^*, \downarrow_{\leq_s} L$ is a finite union of sets of the form*

$$w_0 \Sigma_1^* w_1 \Sigma_2^* w_2 \ldots w_{n-1} \Sigma_n^* w_n,$$

for some $n \geq 0, w_i \in \Sigma^*, \Sigma_i \subseteq \Sigma$, for any $1 \leq i \leq n$.

Proof. Since $\downarrow_{\leq_s} L$ is regular for any L, we may restrict ourselves to regular languages L; just take $\downarrow_{\leq_s} L$ instead of L and use the fact that the operator $\downarrow_{\leq_s}$ is idempotent.

The proof now continues by induction on the regular expression defining L. The basis of the induction, for $L = \emptyset$ or L is a letter, is clear. Suppose then that the statement holds true for L defined be any of the regular expressions α and β and let us prove it for L defined by $(\alpha + \beta)$, $(\alpha\beta)$ or α^*.

The cases of union and catenation are clear. (The catenation of two sets of the form in the statement has the same form.) For α^*, if $\mathrm{alph}(\alpha) \subseteq \Sigma$ denotes the set of letters appearing in α, then the required representation for α^* is just $(\mathrm{alph}(\alpha))^*$.

□

We can state now the connection result already announced.

Theorem 6.6 *If $\leq$ is a well quasi order on Σ^* which is an extension of $\leq_s$, then any language $L \subseteq \Sigma^*$ is confluent w.r.t. $\leq$ in generalized sense.*

Proof. Take $L \subseteq \Sigma^*$. By Lemma 6.5, the down-set $\downarrow_{\leq_s} L$ is a finite union of sets of the form $R = w_0 \Sigma_1^* w_1 \Sigma_2^* w_2 \ldots w_{n-1} \Sigma_n^* w_n$, using the notations there. If $x, y \in R$, then $x = w_0 u_1 w_1 \ldots u_n w_n$, $y = w_0 v_1 w_1 \ldots v_n w_n$, for some

$u_i, v_i \in \Sigma_i^*, 1 \leq i \leq n$. Consider the word $z = w_0 u_1 v_1 w_1 \ldots u_n v_n w_n \in R$. Obviuosly $x \leq_s z, y \leq_s z$, thus R is confluent w.r.t. $\leq_s$. Therefore $\downarrow_{\leq_s} L$ is confluent w.r.t. $\leq_s$ in generalized sense. By Theorem 6.1, we obtain that also L is confluent w.r.t. $\leq_s$ in generalized sense and put

$$L = \bigcup_{i=1}^{k} L_i,$$

for some $k \geq 1, L_i \subseteq \Sigma^*, L_i$ confluent w.r.t. $\leq_s$, for any $1 \leq i \leq k$. Since $\leq$ is an extension of $\leq_s$, it follows that L_i is confluent w.r.t. $\leq$, for any $1 \leq i \leq k$. Therefore, L is confluent w.r.t. $\leq$ in generalized sense, as claimed.

□

Using Theorem 3.1 and Corollary 3.2, we get the following corollary of Theorem 6.6.

Corollary 6.7 *If $\leq \in \{\leq_s, \leq_{\mathcal{P}}\}$, then any language is confluent w.r.t. $\leq$ in generalized sense.*

We end this section with the following historical observation. It is quite surprising that the regular languages of the particular type appearing in Lemma 5.2 and Lemma 6.5 were investigated already in the 60's. Indeed, the terms of the union appearing in Lemma 6.5 are called "quasi-uniform events" in [Ro], where several decision problems are settled, for instance, the problem of an arbitrarily given regular language being a quasi-uniform event. References to further work in this area, in particular by Michael Yoeli, are given in [Ro] and [Sa2]. In general, languages of low loop complexity such as quasi-uniform events were important in the theory of switching circuits. The work of Schützenberger, [DS], showed that the loop complexity can become arbitrarily high.

7 Effective regularity of down-sets

For any language L, the down-set $\downarrow_{\leq_s^{-1}} L$ is regular, see Theorem 3.3. Van Leeuwen [vL] proved a general result concerning the effective regularity of down-sets $\downarrow_{\leq_s^{-1}} L$, see Theorem 3.6. In this section, we give a generalization of this result for well quasi orders with certain properties.

We remind that for any monotone well quasi order $\leq$ on Σ^*, the down-set $\downarrow_{\leq^{-1}} L$ is regular, for any language $L \subseteq \Sigma^*$, see Theorem 3.4.

Let us denote the length quasi order on Σ^* by $\leq_l$; for $u, v \in \Sigma^*$,

$$u \leq_l v \text{ iff } |u| \leq |v|.$$

Theorem 7.1 *Let $\leq$ be a monotone well quasi order on Σ^* such that $\leq \subseteq \leq_l$ and, for any $u, v \in \Sigma^*$, if $u \leq v$ and $|u| = |v|$, then also $v \leq u$. Suppose that, for any finite set F, the down-set $\downarrow_{\leq^{-1}} F$ is effectively regular. If $\mathcal{L}$ is a family of languages effectively closed under intersection with regular sets, then the regularity of the down-sets $\downarrow_{\leq^{-1}} L$ is effective for languages L in $\mathcal{L}$ if and only if the emptiness problem is decidable in $\mathcal{L}$.*

Proof. Suppose first that, for any $L \in \mathcal{L}$, the down-set $\downarrow_{\leq^{-1}} L$ is effectively regular. By the definition of $\downarrow_{\leq^{-1}} L$, it follows that if L is empty, then $\downarrow_{\leq^{-1}} L$ is empty. Since $\leq^{-1}$ is reflexive, $L \subseteq \downarrow_{\leq^{-1}} L$, so, if $\downarrow_{\leq^{-1}} L$ is empty, then L is empty as well. Consequently, L and $\downarrow_{\leq^{-1}} L$ are empty at the same time. The decidability of the emptiness problem in $\mathcal{L}$ follows now from its decidability for regular languages.

Conversely, suppose that the emptiness problem is decidable in $\mathcal{L}$ and let us show that, for any $L \in \mathcal{L}$, the down-set $\downarrow_{\leq^{-1}} L$ is effectively regular.

Consider an arbitrary total order on $\Sigma = \{a_1, a_2, \ldots, a_{\mathrm{card}(\Sigma)}\}$, $a_i < a_j$, for any $1 \leq i < j \leq \mathrm{card}(\Sigma)$, and the total lexicographical order on Σ^*; for $u, v \in \Sigma^*$,

$$\begin{aligned} u \leq_{lex} v \text{ iff either } & |u| < |v|, \text{ or } u = v, \text{ or else} \\ & u = a_{i_1} a_{i_2} \ldots a_{i_{|u|}}, v = a_{i_1} \ldots a_{i_{k-1}} a_{j_k} \ldots a_{j_{|u|}}, \\ & \text{for some } 1 \leq k \leq |u|, 1 \leq i_l, j_l \leq \mathrm{card}(\Sigma), i_k < j_k. \end{aligned}$$

For any language L, define

$$\min\nolimits_{\leq} L = \{w \in L \mid \text{ for any } x \in L, \text{if } x \leq w, \text{then } w \leq x \text{ and } w \leq_{lex} x\}.$$

Claim 1. For any language L, $\min_{\leq} L$ is finite.

Proof of Claim 1. It is enough to show that the set $A = \{|w| \mid w \in \min_{\leq} L\}$ is finite. In order to prove this, consider, for any $n \in A$, $w_n \in \min_{\leq} L, |w_n| = n$, and take the set $B = \{w_n \mid n \in A\}$. If B is finite, then A is finite too. Suppose that, for some $n, m \in A, n \neq m$, we have $w_n \leq w_m$. Then, as $\leq \subseteq \leq_l$, we have $n \leq m$ and so $n < m$. From $w_m \in \min_{\leq} L$ we get $w_m \leq w_n$. It follows as above that $m < n$, a contradiction. Consequently, B is an antichain of $\leq$. As $\leq$ is a well quasi order, B is finite, thus A is finite.

Claim 2. For any language L, $\downarrow_{\leq^{-1}} (\min_{\leq} L) = \downarrow_{\leq^{-1}} L$. Moreover, if $L \neq \emptyset$, then, for any $R \subset \min_{\leq} L$, $\downarrow_{\leq^{-1}} R \subset \downarrow_{\leq^{-1}} L$.

Proof of Claim 2. Since $\min_{\leq} L \subseteq L$, it follows that $\downarrow_{\leq^{-1}} (\min_{\leq} L) \subseteq \downarrow_{\leq^{-1}} L$. To prove the converse inclusion, take $w_0 \in L$. Since $\leq$ is well founded, we can find $w_1 \in L$ such that $w_1 \leq w_0$ and, for any $w \in L$, $w \leq w_1$ implies $w_1 \leq w$. If, for some $w \in L$, $w \leq w_1$ and $w_1 \leq w$, then, by hypothesis,

$|w| = |w_1|$. Hence, we can find $w_2 \in L$ such that $w_2 \leq w_1, w_1 \leq w_2$, and, for any $w \in L$ with $w \leq w_1$ and $w_1 \leq w$, we have $w_2 \leq_{lex} w$. Therefore $w_2 \in \min_{\leq} L$ and, since $w_0 \leq^{-1} w_2$, it follows that $w_0 \in \downarrow_{\leq^{-1}} (\min_{\leq} L)$. The equality is proved.

Take now $R \subset \min_{\leq} L$ and $w \in \min_{\leq} L - R$. (Since L is nonempty, $\min_{\leq} L$ is nonempty too.) We claim that $w \notin \downarrow_{\leq^{-1}} R$. Indeed, if $w \in \downarrow_{\leq^{-1}} R$, then $x \leq w$, for some $x \in R$. As $w \in \min_{\leq} L$, it follows that $w \leq x$ and $w \leq_{lex} x$. As also $x \in \min_{\leq} L$, we get $x \leq_{lex} w$, thus $x = w$ and so $w \in R$, a contradiction. The claim is proved.

We now give an algorithm to compute a regular expression for $\downarrow_{\leq^{-1}} L$.

Algorithm 7.2
Input: a language $L \in \mathcal{L}, L \subseteq \Sigma^*$.
Output: a regular expression representing the regular language $\downarrow_{\leq^{-1}} L$.

1. Consider Σ^* totally ordered lexicographically and put $\Sigma^* = \{w_1, w_2, w_3, \ldots\}$, such that $w_i \leq_{lex} w_{i+1}$, for any $i \geq 1$;

2. Initialize $\mathrm{MIN_L} \longleftarrow \emptyset$

[the variable $\mathrm{MIN_L}$will reach, at the end of the algorithm, the value $\min_{\leq} L$];

3. $i \longleftarrow 1$

[start with w_1];

4. If $L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L})) \neq \emptyset$, then go to step 5, else go to step 6;

[When $L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L})) = \emptyset$, that is, $L \subseteq \downarrow_{\leq^{-1}} (\mathrm{MIN_L})$, the algorithm has to stop because the value of $\min_{\leq} L$ has been found. Indeed, as it will be seen at step 5, always $\mathrm{MIN_L} \subseteq \min_{\leq} L$ and so, by Claim 2, $L \subseteq \downarrow_{\leq^{-1}} \mathrm{MIN_L}$ if and only if $\mathrm{MIN_L} = \min_{\leq} L$. Otherwise, the algorithm goes to step 5 attempting to include w_i in $\mathrm{MIN_L}$. By hypothesis, $\downarrow_{\leq^{-1}} (\mathrm{MIN_L})$ is effectively regular since, by Claim 1, $\min_{\leq} L$ is finite. Then $\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L})$ is effectively regular too. Thus $L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L}))$ is effectively in $\mathcal{L}$ and its emptiness can be decided.]

5. If $(L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L}))) \cap \{w_i\} \neq \emptyset$, then $\mathrm{MIN_L} \longleftarrow \mathrm{MIN_L} \cup \{w_i\}, i \longleftarrow i + 1$, and go to step 4, else $i \longleftarrow i + 1$, and go to step 5.

[Notice first that, as seen above, $L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L}))$ is effectively in $\mathcal{L}$, so $(L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L}))) \cap \{w_i\}$ is effectively in $\mathcal{L}$ too, as $\{w_i\}$ is regular. Therefore, the emptiness of the language $(L \cap (\Sigma^* - \downarrow_{\leq^{-1}} (\mathrm{MIN_L}))) \cap \{w_i\}$ can be decided. If it is not empty, then $w_i \in L$ and $w_i \notin \downarrow_{\leq^{-1}} (\mathrm{MIN_L})$. Now, if $w_j \leq w_i$, for some $w_j \in L$, then $|w_j| \leq |w_i|$. If $|w_j| < |w_i|$, then $j \leq i - 1$. Hence, w_j has been considered at step 5 before, so either $w_j \in \mathrm{MIN_L}$ or

$w_k \leq w_j$, for some $1 \leq k \leq j-2$ such that $w_k \in$ MIN$_\mathrm{L}$. In both cases we find $w_l \in$ MIN$_\mathrm{L}$ such that $w_l \leq w_i$, thus $w_i \in \downarrow_{\leq^{-1}}$ (MIN$_\mathrm{L}$), a contradiction. Therefore $|w_j| = |w_i|$ and so, by hypothesis, $w_i \leq w_j$. Consider $j_0 = \min\{j \mid w_j \in L, w_j \leq w_i, |w_j| = |w_i|\}$. Then $j_0 \leq i$. If $j_0 < i$, then $w_{j_0} \in$ MIN$_\mathrm{L}$ and so $w_i \in \downarrow_{\leq^{-1}}$ MIN$_\mathrm{L}$, a contradiction. Thus $i = j_0$ and $w_i \leq_{lex} w_j$, for any $w_j \in L, w_j \leq w_i, |w_j| = |w_i|$. Consequently $w_i \in \min_\leq L$ and it is correctly added to MIN$_\mathrm{L}$. On the other hand, if $(L \cap (\Sigma^* - \downarrow_{\leq^{-1}}(\mathrm{MIN_L}))) \cap \{w_i\} = \emptyset$, then, as $L \cap (\Sigma^* - \downarrow_{\leq^{-1}}(\mathrm{MIN_L}) \neq \emptyset$ from step 4, either $w_i \notin L$ or $w_j \leq w_i$, for some $j < i$ such that $w_j \in$ MIN$_\mathrm{L} \subseteq \min_\leq L$. In both cases, w_i is correctly skipped and w_{i+1} is tested at step 5. If w_i is added, then, before testing w_{i+1}, the algorithm checks at step 4 whether there is any word to be added to MIN$_\mathrm{L}$.]

6. Output a regular expression for $\downarrow_{\leq^{-1}}(\min_\leq L)$.

[As, by Claim 1, $\min_\leq L$ is finite, the algorithm will certainly reach this step and end. By hypothesis, it is possible to find effectively a regular expression for $\downarrow_{\leq^{-1}}(\mathrm{MIN_L})$. Step 6 is reached when $L \cap (\Sigma^* - \downarrow_{\leq^{-1}}(\mathrm{MIN_L})) = \emptyset$, so $L \subseteq \downarrow_{\leq^{-1}}(\mathrm{MIN_L})$. It follows by Claim 2 that MIN$_\mathrm{L}$ has reached the value MIN$_\mathrm{L} = \min_\leq L$ and the regular expression for $\downarrow_{\leq^{-1}}(\min_\leq L)$ at output is also a regular expression for $\downarrow_{\leq^{-1}} L$, as required.]

□

As an application of our general result in Theorem 7.1, we give the next corollary.

Corollary 7.3 *For any context-free language L, the down-set $\downarrow_{\leq_\mathcal{P}^{-1}} L$ is effectively regular.*

Proof. The quasi order $\leq_\mathcal{P}$ is a monotone well quasi order on Σ^* and, obviously, $\leq_\mathcal{P} \subseteq \leq_l$. Also, if $u \leq_\mathcal{P} v$ and $|u| = |v|$, then u and v have the same Parikh vector, so also $v \leq u$.

For a finite set $F = \{w_1, w_2, \ldots, w_n\} \subset \Sigma^*$, if $\lambda \in F$, then $\downarrow_{\leq_\mathcal{P}^{-1}} F = \Sigma^*$. Otherwise, if $w_i = w_{i,1}w_{i,2}\ldots w_{i,n_i}, w_{i,j} \in \Sigma, 1 \leq j \leq n_i, 1 \leq i \leq n$, then

$$\downarrow_{\leq_\mathcal{P}^{-1}} F = \bigcup_{i=1}^{n} \bigcup_{\pi \in \mathcal{S}_{n_i}} \Sigma^* w_{i,\pi(1)}\Sigma^* w_{i,\pi(2)}\Sigma^* \ldots \Sigma^* w_{i,\pi(n_i)}\Sigma^*,$$

so it is effectively regular.

Since the family of context-free languages fulfills the conditions in Theorem 7.1, the result is proved.

□

References

[CK] C. Choffrut, J. Karhumäki, Combinatorics of Words, in: G. Rozenberg, A Salomaa, eds., *Handbook of Formal Languages* (Springer-Verlag, Berlin, Heidelberg, 1997).

[DS] F. Dejean, M. P. Schützenberger, On a question of Eggan, *Information and Control*, **9** (1966) 23–25.

[dLV] A. de Luca, S. Varricchio, Unavoidable Regularity and Finiteness Conditions, in: G. Rozenberg, A Salomaa, eds., *Handbook of Formal Languages* (Springer-Verlag, Berlin, Heidelberg, 1997).

[ER] P. Erdös, R. Rado, Sets having divisor property, Solution to problem 4358, *Amer. Math. Monthly* **59** (1952) 255–257.

[Ha] L. H. Haines, On free monoids partially ordered by embedding, *J. Comb. Theory* **6** (1969) 94–98.

[HI] T. Harju, L. Ilie, On quasi orders of words and the confluence property, manuscript, 1996.

[Hi] G. Higman, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* (3) **2** (1952) 326–336.

[HU] J. E. Hopcroft, J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading, Mass., 1979).

[Kr1] J. B. Kruskal, Ph.D. Thesis, The theory of well-partially-ordered sets, Princeton, 1954.

[Kr2] J. B. Kruskal, The theory of well-quasi-ordering; a frequently discovered concept, *J. Comb. Theory* (A) **13** (1972) 297–305.

[Lo] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, Mass., 1983).

[Ro] G. Rozenberg, Decision problems for quasi-uniform events, *Bull. Acad. Polon. Sci.*, Ser. sci. math., astr., phys. Vol. XV (1967) 745–752.

[Sa1] A. Salomaa, *Formal Languages* (Academic Press, New York, 1973).

[Sa2] A. Salomaa, *Theory of Automata* (Pergamon Press, Oxford, 1969).

[vL] J. van Leeuwen, Effective constructions in well-partially-ordered free monoids, *Discrete Mathematics* **21** (1978) 237–252.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**